

1 ABSTRACT

2 A method for the joint optimization of language model performance and
3 size is presented comprising developing a language model from a tuning set of
4 information, segmenting at least a subset of a received textual corpus and
5 calculating a perplexity value for each segment and refining the language model
6 with one or more segments of the received corpus based, at least in part, on the
7 calculated perplexity value for the one or more segments.

8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25